

Fast Millimeter Wave Beam Alignment

Haitham Hassanieh*
UIUC

Omid Abari*
University of Waterloo

Michael Rodriguez
MIT

Mohammed Abdelghany
UCSB

Dina Katabi
MIT

Piotr Indyk
MIT

*Co-Primary Authors

ABSTRACT

There is much interest in integrating millimeter wave radios (mmWave) into wireless LANs and 5G cellular networks to benefit from their multi-GHz of available spectrum. Yet, unlike existing technologies, e.g., WiFi, mmWave radios require highly directional antennas. Since the antennas have pencil-beams, the transmitter and receiver need to align their beams before they can communicate. Existing systems scan the space to find the best alignment. Such a process has been shown to introduce up to seconds of delay, and is unsuitable for wireless networks where an access point has to quickly switch between users and accommodate mobile clients.

This paper presents Agile-Link, a new protocol that can find the best mmWave beam alignment without scanning the space. Given all possible directions for setting the antenna beam, Agile-Link provably finds the optimal direction in logarithmic number of measurements. Further, Agile-Link works within the existing 802.11ad standard for mmWave LAN, and can support both clients and access points. We have implemented Agile-Link in a mmWave radio and evaluated it empirically. Our results show that it reduces beam alignment delay by orders of magnitude. In particular, for highly directional mmWave devices operating under 802.11ad, the delay drops from over a second to 2.5 ms.

CCS CONCEPTS

• **Networks** → **Network algorithms**; **Wireless access networks**; *Network protocols*; *Physical links*; *Mobile networks*;

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGCOMM '18, August 20–25, 2018, Budapest, Hungary

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-5567-4/18/08...\$15.00

<https://doi.org/10.1145/3230543.3230581>

KEYWORDS

Millimeter Wave, Sparse Recovery, 5G, Beam Alignment

1 INTRODUCTION

The ever-increasing demands for mobile and wireless data have placed a huge strain on wireless networks [10, 43]. Millimeter wave (mmWave) frequency bands address this problem by offering multi-GHz of unlicensed bandwidth, 200× more than the bandwidth allocated to today’s WiFi and cellular networks [30, 34]. They range from the 24GHz ISM band all the way to hundreds of GHz [1]. They are expected to play a central role in dealing with increased multimedia traffic, the introduction of new high data-rate applications such as virtual reality, and the anticipated surge in IoT wireless devices [2, 4]. This role has been cemented with new standards that incorporate mmWave technologies into 5G networks [25, 34], and 802.11 wireless LAN [22].

Millimeter wave radios however do not play well with mobile devices or dynamic environments, a key challenge that has been emphasized in the standards [22, 28]. Specifically, mmWave signals attenuate quickly with distance; hence they need to use highly directional antennas to focus their power on the receiver. Luckily, due to their small wavelength (millimeter scale), it is possible to pack hundreds or thousands of antennas in a small area, creating an array with many antennas, and hence a very narrow beam, as shown in Figs. 1(a) and (b). Yet, since the beam is very narrow, communication is possible only when the transmitter’s and receiver’s beams are well aligned. Current solutions for aligning the beams scan the entire space, trying various beam alignments until they find the best one. This process can take up to several seconds [39, 47]. Such a long delay makes the deployment of mmWave links infeasible in wireless networks, where the access point has to keep realigning its beam to switch between users and accommodate mobile clients.

To understand the problem, consider a phased array with many antennas. Its beam-width can be a few degrees or even smaller. The naive approach to finding the best alignment would have the transmitter and receiver scan the 3D space with their beams to find the direction of maximum power, as shown in Fig. 1(b). The receiver has to repeat the scan for

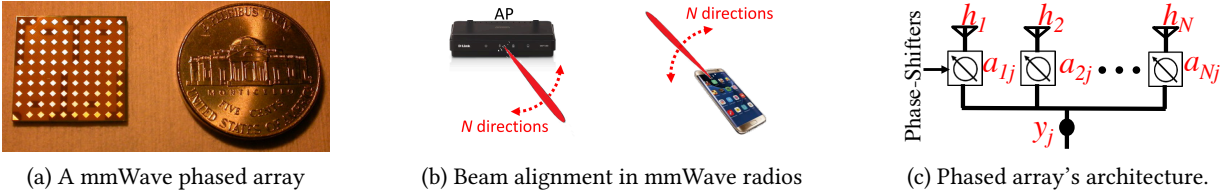


Figure 1: mmWave Communication (a) An example of mmWave phased arrays where hundreds of antennas are packed in a small area [21]. (b) mmWave radios need to align their beams to establish a communication link. (c) mmWave phased arrays use a set of phase shifters to steer the beam.

each choice of beam direction on the transmitter side. Thus, the complexity of this exhaustive search is $O(N^2)$, where N is the number of possible beam directions. To speed up the search, the 802.11ad standard decouples the steering at the transmitter and receiver. In particular, the transmitter starts with a quasi-omnidirectional beam, while the receiver scans the space for the best beam direction. The process is then reversed to have the transmitter scan the space while keeping the receiver quasi-omnidirectional [22, 46] (see §6.1 for the details). This approach reduces the search complexity to $O(N)$. Still, for a beam of a few degrees, the delay can be hundreds of milliseconds to seconds [39, 47], which would easily stall realtime applications.

But, can one identify the best alignment without scanning the space of all possible signal directions? In principle, "yes". There is much past work that shows that mmWave signals travel along a small number of paths, e.g., 2 or 3 paths [6, 34]. This means that the space of possible signal directions is sparse. One would hope to use the sparse recovery theory to estimate the direction of the best alignment using a logarithmic number of measurements [9, 11, 18], hence avoiding excessive delays.

Problem Formalization: The objective of beam alignment is to measure the signal power along each spatial direction. Let \mathbf{x} be an N -element vector that denotes the signal along various spatial directions. Since in practice the signal arrives only along few directions K , we can say that \mathbf{x} is K -sparse. Let, h_i be the signal at the i^{th} antenna, as shown in Fig. 1(c). Based on the standard antenna array equation [44], we can write $\mathbf{h} = \mathbf{F}'\mathbf{x}$, where \mathbf{F}' is the inverse Fourier transform matrix. We can steer the antenna beam by applying a phase shift to each antenna $a_i = e^{-j2\pi\phi_i}$ (see Fig. 1(c)). For each setting of the phase shifters, we can measure the received power as $y_j = |\mathbf{a}_j\mathbf{F}'\mathbf{x}|$, where the notation $|\cdot|$ refers to the magnitude of the signal and \mathbf{a}_j is a vector whose elements are the applied phase shifts. Note that knowing the magnitude is the same as knowing the power since power is magnitude squared. In 802.11ad, each measurement corresponds to sending a special frame. Each time a frame is sent, the signal incurs a random phase due to the Carrier Frequency Offset (CFO) between the transmitter and receiver [22, 32] (see §4.1 for details).

Thus, one cannot compare the phase of two measurements; only the magnitude of the measurements is relevant. Since we want to know the power along each spatial direction, the problem can be formulated as:

$$\text{estimate each } |x_i|, \text{ given measurements } y_j = |\mathbf{a}_j\mathbf{F}'\mathbf{x}|.$$

Of course, one way to solve this problem is to use N measurements, each time setting \mathbf{a} to one row of the Fourier matrix. This corresponds to measuring one direction every time, as proposed in 802.11ad. Alternatively, one could leverage that the vector \mathbf{x} is sparse, and hope to solve the problem in a logarithmic number of measurements. Unfortunately, however, using off-the-shelf algorithms like compressive sensing or the sparse FFT does not work since neither of them deal with the scenario where the measurements return the magnitude of the complex signal, i.e., the presence of the $|\cdot|$ term [9, 11, 18]. To the best of our knowledge, no algorithm with provable logarithmic guarantees exist for this problem.

Our Design & Analysis: This paper introduces Agile-Link, the first solution that provably finds the best beam alignment in a logarithmic number of measurements. At a high-level, it works as follows: Instead of creating a narrow beam and sampling the power along one spatial direction each time, Agile-Link manipulates the phase shifters to create multi-armed beams, which can sample multiple spatial directions simultaneously (see Fig 2(a)). Since a multi-armed beam combines the power along multiple directions, the receiver cannot immediately tell which direction has produced the resulting power. Agile-Link however uses a combination of randomized multi-armed beams, which together provide enough information to identify the signal power along all spatial directions. We formally analyze Agile-Link and prove that it can deliver the best alignment in $O(K \log N)$ measurements, where K is the number of paths traveled by the signal. Since K is typically 2 or 3 paths [6, 34], Agile-Link significantly reduces the beam alignment delay.

Agile-Link has additional important features. First, Agile-Link is compatible with the 802.11ad protocol, i.e., a Agile-Link device can work with a non-Agile-Link device to find the best alignment while using the 802.11ad protocol. In this

case, the Agile-Link device finds the best alignment on its side in a logarithmic number of measurements whereas the traditional 802.11ad device takes a linear number of measurements. Second, by the results of [7, 31], our measurement complexity of $O(K \log N)$ is asymptotically *optimal* for small K –i.e., it cannot be further reduced.

Implementation & Empirical Results: We have also implemented Agile-Link and evaluated it using mmWave radios, each equipped with a phased array that has 8 antennas. We also use simulations to explore its scaling behavior to large arrays with hundreds of antennas, which are expected in the future [8]. We compare Agile-Link with two baselines: an exhaustive scan of the space to find the best beams, and the quasi-omnidirectional search proposed in the 802.11ad standard. Our evaluation reveals the following findings.

- In comparison with the exhaustive search, Agile-Link reduces the search time by one to three orders of magnitude for array sizes that range from 8 antennas to 256 antennas. In comparison to the quasi-omni directional search, Agile-Link reduces the number of measurements by $1.5\times$ to $16.4\times$ for the same range of array sizes. In particular, for large arrays with 256 antennas, Agile-Link reduces the alignment delay from over a second to 2.5ms.
- The quasi-omnidirectional search yields poor performance in scenarios with multi-path effects. This is because using a quasi-omnidirectional antenna allows the signals along different paths to combine destructively, which yields low power and prevents accurate detection of the best signal direction. Further, due to imperfections in the quasi-omni directional patterns [20, 27], some paths can get attenuated and hence this approach can choose the wrong direction to align its beam. In contrast, Agile-Link performs well both in single path and multipath scenarios.

2 RELATED WORK

(a) mmWave Solutions & Research: Research on fast beam alignment for mmWave can be divided into two classes: empirical and simulation-based. Past empirical work has demonstrated the large delays incurred during beam alignment [39, 47]. It also proposed failover protocols that switch to the next best beam when the current beam becomes blocked [16, 40]. This approach however assumes that the signal propagation paths are known a priori and hence one can quickly switch to a failover direction.

Much of the previous work on fast beam alignment is simulation-based. Most of this work proposes enhancements to the standard that impose a form of hierarchy to speed up the search [26, 41, 45]. Hierarchical algorithms, however, are not robust to multipath. This is due to the fact that different paths can combine destructively and cancel each other at

any level of the hierarchy. In §3(b), we present an example that illustrate this case. Furthermore, in practice, hierarchical search requires feedback from the client to guide the access point at every stage of the hierarchy, which incurs significant protocol delay [35].

Our work is closest to past work that leverages compressive sensing to speed up the search for best beam alignment. The work in [5, 12] requires a more complex architecture with a quadratic number of phase shifters, and multiple transmit receive chains (typically 10 to 15 [5]). The best known results for such complex architecture can guarantee a logarithmic number of measurements *only* for scenarios with no multi-path (K is strictly 1), and only for the average case error [5]. The work in [35] also uses compressive sensing to speed up the search but it does not provide any theoretical guarantees. In comparison, Agile-Link can provably find the best alignment in a logarithmic number of measurements and its guarantees apply to the worst case behavior. In §6.5, we empirically compare the performance of Agile-Link with the compressive sensing scheme proposed in [35].

Some past theoretical work applies the standard compressive sensing and assumes it can correctly obtain the phase of the measurements [13, 33]. This approach does not work with practical 802.11ad or cellular radios because it ignores CFO (Carrier Frequency Offset) which corrupts the phase of the measurements [32, 35] as we describe in §4.1.

Finally, some companies [8, 36, 38] offer mmWave systems but they take a long time to steer the beam or require complex hardware, making them unsuitable for mobile clients [39, 47]. Also, some research on mmWave focuses on point-to-point links for Data Centers [17] or cellular and WiFi applications [16, 29, 39, 40, 47]. These works typically use a horn antenna to direct the beam, which requires mechanical steering and is unsuitable for mobile links.

(b) Sparse Recovery Theory: The theoretical problem we consider falls under “sparse phase retrieval” [23, 24]. Generally, the goal is to recover an approximation of a K -sparse vector, \mathbf{x} , from M measurements of the form $|\mathbf{B}\mathbf{x}|$. The presence of the absolute value is what makes this problem different from the usual compressive sensing [9, 11] and sparse FFT [14, 15, 18, 19]. In our context, we have an extra restriction that the matrix \mathbf{B} is of the form $\mathbf{A}\mathbf{F}'$, where \mathbf{F}' is the inverse Fourier transform and all entries in \mathbf{A} have unit magnitude, i.e., $|a_{ij}| = 1$ for all i, j . To the best of our knowledge, this form has not been considered before.

Some of our proof techniques are inspired by past work on sparse FFT, particularly the work in [14, 15] which used boxcars filters for sparse Fourier transform algorithms. However, the technical development of our proofs is different due to the leakage between multiple beam arms, which requires extra layers of randomization. Furthermore, our results and

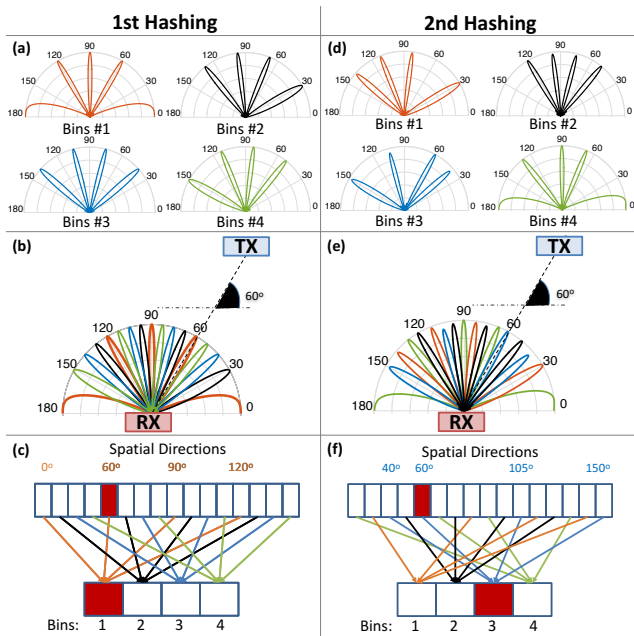


Figure 2: Illustrative example of Agile-Link's algorithm

problem are different due to the restrictions on our measurements that do not exist in sparse FFT.

(c) Massive MIMO Beamforming: Massive MIMO [37] has many antennas, but the signal from each antenna is received and manipulated independently. In contrast, a mmWave phased array receives only the combined signal from its antennas. Thus, beamforming techniques in standard and massive MIMO do not apply to mmWave phased array.

3 ILLUSTRATIVE EXAMPLE

(a) Single Path Example: We start by explaining the intuition of our algorithm. Consider a case where the transmitter is at a 60° angle with respect to the receiver. For clarity, assume a 2D setting. The same argument can be extended to 3D. We would like the receiver to detect that the best alignment is along the 60° angle without scanning the space.

Agile-Link avoids the need to scan all spatial directions by using multi-armed beams, which simultaneously sample the signal along multiple directions. Say for example, that there are 16 possible directions in space, i.e., $N = 16$. Agile-Link can sample all of these directions using 4 multi-armed beams, each covering $N/4 = 4$ directions in space. Fig. 2(a) shows four such multi-armed beams, and Fig. 2(b) shows how together they cover the whole space of directions. Such set of multi-armed beams operates like a hash function, where $N = 16$ directions are hashed into 4 bins, and each bin covers $N/4 = 4$ distinct directions. The value of the bin represents the combination of the signals that hash into it. For example, if the signal is coming along the 60° direction and 60° hashes

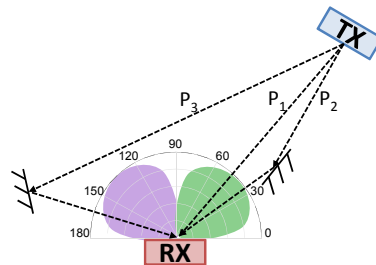


Figure 3: Example of hierarchical search with multipath.

to bin number 1, then only bin one will have energy whereas the other bins will have no energy, as shown in Fig. 2(c). Thus, one can ignore directions that hash to bins 2,3, and 4, and focus only on directions that hash to bin 1. This significantly reduces the search space to the directions that hash to the first bin (i.e., the first multi-armed beam).

At this stage, we know that the signal could have come from the directions covered by the first bin i.e., 0° , 60° , 90° and 120° . But we do not know which among them is the correct direction. Thus, we change the hash function and try again. To do so, we use a second set of multi-armed beams which together hash the whole space of directions into a set of bins. The hash however is randomized with respect to the previous hash so that directions that got hashed together are unlikely to hash together again. Figs. 2(d,e,f) show an example of hashing the spatial directions into bins after randomizing the multi-armed beams and hence randomizing which directions map into which bins. The first bin now collects energy along 30° , 80° , 110° and 140° . Since the signal is arriving along 60° , it will be captured by the third bin which is represented in blue in Fig. 2(f). Hence, in this second hashing, only the energy of the third bin will be large. This suggests that the signal arrived along one of the directions that mapped to the third bin which in this case are 40° , 60° , 105° and 150° . Since the 60° direction is the only common candidate from both the first hashing and the second hashing, Agile-Link picks it as the direction of the signal. Thus, Agile-Link is able to find the correct direction from which the transmitter's signal arrives without having to scan all possible directions.¹

(b) Multipath Example: But, what if the signal arrives along multiple paths? For example, in Fig. 3, the signal arrives along three directions: p_1 , p_2 and p_3 . Paths p_1 and p_2 are higher power than p_3 because they are significantly shorter; and the best beam alignment is along the direction of p_1 which delivers the highest power.

¹While the above example describes the algorithm for recovering one of N possible direction, Agile-Link can recover the optimal direction at a finer resolution by taking a continuous weight over possible choice of directions as described in detail in §4.

Let us compare two beam alignment algorithms: Hierarchical Search and Agile-Link. As explained in §2, multiple past proposals use a hierarchy of beams to speed up the search for the best alignment [26, 41, 45]. Proposals differ in the details, but they use wide beams at the top of the hierarchy – e.g., they might start with two wide beams, check which beam returns more power, then zoom in on the part of the space that returns more power and use narrower beams to explore it. Like Agile-Link, Hierarchical Search needs only a logarithmic number of measurements. The problem is that using wide beams means nearby signal directions would collide within the same beam. Since RF signals are waves, the colliding signals can combine destructively to cancel each other’s power. In our example, p_1 and p_2 have close directions and thus will collide in the wide beam. Say that p_1 and p_2 have opposite phases, they will cancel each other’s power, making it look as if the signal from the left-half of the space (i.e., directions 90° to 180°) had more power. As a result, Hierarchical Search will zoom in on those directions and explore them with narrower beams until it finds path p_3 , which is the worse alignment out of the three signal directions. Note that this failure mode can be quite common and does not require the phases of p_1 and p_2 to be exact opposite; it is sufficient that they point away from each other so that the resulting power degradation makes it look as if p_3 were higher power.

In contrast, Agile-Link does not use wide beams, where nearby directions are bound to collide. Agile-Link uses multi-armed beams where the arms in each beam can be randomized. Even if two paths collide in one multi-armed beam, our randomization ensures that the same two paths will not collide in other multi-armed beams, as formalized in §4.2. Hence, Agile-Link can recover all possible paths and then pick the best path to align the beam. At a high level, the system operates as multiple randomized hash functions, which can be provably resolved to recover all signal directions even in the presence of multipath.

4 AGILE-LINK

This section describes Agile-Link in detail. For clarity, we describe the algorithm assuming only the receiver has an antenna array whereas the transmitter is omni-directional. The extension to the case where both transmitter and receiver have antenna arrays is described in §4.4.

4.1 Problem Statement

Recall that the problem is defined as follows: Let \mathbf{x} be a K -sparse N -element complex vector that denotes the signal along various spatial directions. The objective is to estimate the power (i.e., magnitude) of the signal along each direction, $|x_i|$, using a small number of measurements of the form

$y_j = |\mathbf{a}_j \mathbf{F}' \mathbf{x}|$, where \mathbf{F}' is the inverse Fourier matrix, and \mathbf{a}_j is a vector of phase shifts, $|a_{ij}| = 1$, that are under our control.

Before describing our solution to this problem, it is important to understand why one can only measure the magnitude, not the phase. Every measurement involves sending a frame from transmitter to receiver. Since the oscillators on the transmitter and receiver always experience some CFO (Carrier Frequency Offset) [32], the signal of each frame incurs an additional unknown phase shift. Further, this phase shift changes across frames. Correcting the CFO across measurement frames is not supported in the 802.11ad standard [22]. Furthermore, such a correction will be very hard due to the high frequencies of mmWave. For example, a small offset of 10 ppm at such frequencies can cause a large phase misalignment in less than hundred nanoseconds.

4.2 Agile-Link’s Algorithm

Agile-Link works in two stages. First, it randomly hashes the space into bins (using multi-armed beams) such that each bin collects power from a range of directions. Second, it uses a voting mechanism to recover the directions that have the power. Below, we describe these two stages in detail.

1. Hashing Spatial Directions into Bins

Agile-Link hashes the signal along various directions to bins using multi-armed beams. Let us refer to the arms in each multi-arm beam as the *sub-beams*. Let R be the number of sub-beams in each beam, B the number of bins in each hash function, and L the number of hash functions. Each setting of the phase shifter vector, \mathbf{a} , creates a different beam pattern and the resulting measurement $y = |\mathbf{a} \mathbf{F}' \mathbf{x}|$ will correspond to the power in the directions covered by the beam pattern.

So how do we create good multi-armed beams? It should be clear that, given the structure of the measurements, we can create a beam that points in one direction, s , by setting \mathbf{a} to the s -th row in the Fourier matrix. Thus, to create a multi-armed beam, Agile-Link divides the vector \mathbf{a} into R segments each of length N/R i.e., $\mathbf{a}_{[1:N/R]}, \mathbf{a}_{[N/R+1:2N/R]}, \dots, \mathbf{a}_{[(R-1)N/R:N]}$. Each segment then sets its sub-beam towards a different direction. This is done by setting the segment $\mathbf{a}_{[1+(r-1)N/R:rN/R]}$ to the corresponding segment in the desired row of the Fourier matrix. Formally, if an index i belongs to the r -th segment pointing towards the direction s^r , then $a_i = (\mathbf{F}_{s^r})_i \cdot e^{-j2\pi t_r/N}$, where t_r is a random integer between $0, \dots, N - 1$, and $(\mathbf{F}_{s^r})_i$ is the i -th entry in row s^r in the Fourier matrix. The term $e^{-j2\pi t_r/N}$ results in a phase shift of the sub-beam without changing its direction.²

Due to the properties of the Fourier transform, the sub-beam created by each segment will be wider than a single

²An earlier version of this algorithm appeared in a short workshop paper which is not cited due to anonymity. The algorithm we present in this paper is more efficient and can provide proven guarantees.

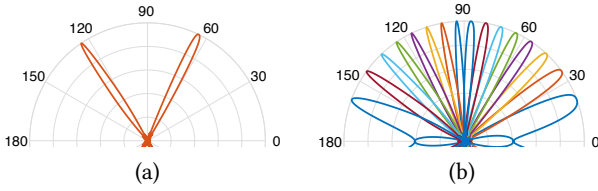


Figure 4: Hashing Beam Patterns (a) Agile-Link hashes well spread directions into a bin; (b) All the bins of Agile-Link's hash function where the same color corresponds to the same bin.

beam created by the full array by a factor of R , so each sub-beam covers R adjacent directions. Since there are R such sub-beams, the multi-armed beam created by this setting of the vector \mathbf{a} will cover R^2 directions. Now, if we wish to hash the space of directions into B bins, then each will cover R^2 directions and hence $B = N/R^2$.

But how do we pick the directions of the sub-beams in each multi-armed beam? The best scenario is when the sub-beams in each multi-armed beam are well spaced so that the leakage from their side-lobes is minimized. Fig 4(a) shows an example of well-spaced sub-beams, where we have a multi-armed beam with two sub-beams directed 60° apart. In this case, the beam pattern will hash directions that are 60° apart into the same bin. By shifting the direction, we can then create all different bins in the hash function, as shown in Fig 4(b) where each color corresponds to a bin in the hash function. Formally, the above process is achieved by setting the direction of the r -th segment in bin b to be equal to $s_b^r = Rb + rP$, where $P = N/R$ is the spacing between two sub-beams corresponding to the same bin.

The question that remains is how do we randomize the hashing process to make sure that two large signals are not hashed to the same bin every time. Ideally, we can solve this problem by randomly permuting the entries in of the vector \mathbf{x} . Physically, however we cannot permute \mathbf{x} . Instead, we leverage a nice property of the Fourier transform that says that we can pseudo-randomly permute the *input* of the Fourier transform (i.e., the entries of the vector \mathbf{x}) by pseudo-randomly permuting its output (i.e., the entries of the vector $\mathbf{F}'\mathbf{x}$) [14, 15, 18]. In the simplest setting, suppose that we want to permute the input samples according to: $\mathbf{x}'(f) = \mathbf{x}(\sigma f)$, where σ is chosen at random for an appropriate distribution, and all operations on indices are done modulo N . This is equivalent to permuting the output $\hat{\mathbf{x}} = \mathbf{F}'\mathbf{x}$ according to: $\hat{\mathbf{x}}'(t) = \hat{\mathbf{x}}(\sigma^{-1}t)$, where σ^{-1} is the inverse of σ modulo N . This enables us to “randomize” the positions of large entries in \mathbf{x} by rearranging the entries of $\mathbf{F}'\mathbf{x}$. This result is very useful since we can easily permute $\mathbf{F}'\mathbf{x}$ in our measurements by simply permuting the elements in the vector \mathbf{a} , i.e., by permuting the phase shifts applied to the phase shifters on each antenna.

More generally, let us use the vector $\mathbf{y}_{1 \times B} = |\mathbf{A}_{B \times N} \mathbf{F}'\mathbf{x}|$ to refer to the B measurements performed as part of a hash function, where \mathbf{A} is a matrix of phase shifts. To permute this hash function, we measure $\mathbf{y}_{1 \times B} = |\mathbf{A}\mathbf{P}'\mathbf{F}'\mathbf{x}|$, where we use the term $|\cdot|$ to refer to the magnitude of the individual elements in the vector, and the matrix \mathbf{P}' is a generalized permutation matrix.³ That is, our phase shifter matrix is equal to $\mathbf{A}\mathbf{P}'$ (note that each entry of the latter matrix has a unit magnitude, i.e., it represents proper phase shifts). This is equivalent to measuring $|\mathbf{A}\mathbf{F}'\mathbf{P}\mathbf{x}|$, for the corresponding generalized permutation matrix \mathbf{P} .

2. Recovering the Directions of the Actual Paths

After hashing the spatial directions into bins, Agile-Link discovers the actual directions of the signal using a voting scheme where each bin gives votes to all directions that hash into that bin. After few random hashes, the directions that have energy will collect the largest number of votes which allows Agile-Link to recover them. Unfortunately, directly applying this voting approach does not work well because the side-lobes of the beams create leakage between the bins and hence a strong path in one bin can leak energy into other bins which corrupts the voting process. To overcome this problem, Agile-Link takes into account the leakage between the bins.

Specifically, Agile-Link models the beam patterns (example of which are shown in Fig. 4(a) and Figs. 2(a) and (d),) as a *coverage function* $I(b, \rho, i)$ that indicates the coverage of the direction i by the beam corresponding to the bin b , assuming the indices are permuted by ρ . It is defined as: $I(b, \rho, i) = |\mathbf{a}^b \mathbf{F}'_{\rho(i)}|^2$, where \mathbf{a}^b is the vector defining the settings of the phase shifters corresponding to bin b .

If we hash into B bins, we will have B such patterns and collect B measurements $\mathbf{y}_{1 \times B}$ corresponding to the bins. After taking the magnitude squared of each measurement, we can estimate the energy of the signal coming from direction i as:

$$T(i, \rho) = \sum_{b=0}^{B-1} y_b^2 \times I(b, \rho, i), \quad (1)$$

If the estimate $T(i, \rho)$ exceeds a predefined threshold T then we conclude that there is a signal coming from direction i . Otherwise we conclude that there is no such signal.

4.3 Performance Analysis

A detailed analysis with full proofs is provided in the appendix. Here, we note the main theorems and their implications.

³ \mathbf{P}' is a generalized permutation matrix if each row or column of \mathbf{P}' contains exactly one non-zero entry, and that entry has unit magnitude. We use the matrix \mathbf{P}' defined as in [18]: the i -th column of \mathbf{P}' contains the value $\omega^{a\sigma i}$ in the row $\sigma(i - b)$, where $\omega = e^{2\pi j/N}$ and a, b, σ are randomly chosen parameters. This has the effect of rearranging the vector \mathbf{x} by moving the entry x_i to $x_{\rho(i)}$ for $\rho(i) = \sigma^{-1}i + a$, and multiplying it by $\omega^{bj + \sigma ba}$.

THEOREM 4.1. *Suppose that the vector \mathbf{x} has at most K non-zero entries, and that the energy x_i^2 of each non-zero entry x_i is at least $1/K$. Furthermore, suppose that N is a prime. There exists a setting of parameters T , R and $B = O(K)$, so that for each candidate direction i , we have that:*

- If $x_i \neq 0$ then $T(i, \rho) \geq T$ with probability at least $2/3$
- If $x_i = 0$ then $T(i, \rho) < T$ with probability at least $2/3$

The probability of correctness can be amplified by performing several (L) random hashes and aggregating the results. There are multiple ways of performing the aggregation. A simple approach is to use “hard voting”, i.e., conclude that a signal is coming from a direction i if the signal from that direction has been detected by the majority of hashes. By Chernoff bound, this approach reduces the probability of incorrect detection from $1/3$ to $e^{-C''L}$ for some constant C'' . By letting $L = O(\log N)$, we can compute correct estimates for all *all* indices i , with the probability of failure at most $1/N$. The algorithm uses $BL = O(K \log N)$ measurements and its running time complexity is $O(NK \log N)$.

The algorithm can be also used to provide estimations of the values of $|x_i|^2$'s. The guarantees are provided by the following theorem. Note that no assumptions about the sparsity of \mathbf{x} are required, although the guarantees are meaningful only for x_i 's whose magnitude is large enough. This makes the estimate resilient to the presence of small amounts of noise at all coordinates.

THEOREM 4.2. *Suppose that N is a prime. There exists a setting of parameters R , $B = O(K)$, and a constant $C > 1$ so that for each candidate direction i , we have:*

$$\Pr[|x_i|^2/C - \|\mathbf{x}\|_2^2/K \leq T(i, \rho) \leq C|x_i|^2 + \|\mathbf{x}\|_2^2/K] \geq 2/3$$

Few points are worth noting. First, in practice, we drop the assumption that N is prime. Second, we use soft voting instead of the hard voting. Specifically, we estimate of the strength of the signal along direction i as $S(i) = \prod_{l=1}^L T_l(i, \rho)$, where $T_l(i, \rho)$ is defined as in Equation 1, for the l -th permutation. The soft voting approach uses more information about the measurements than hard voting, and hence its practical performance is better. We extract the significant coefficients of the signal x_i by selecting the indices i with the largest values of $S(i)$.

4.4 Extension of the Model to Both Transmitter and Receiver

Here, we extend our model and algorithm to the case where both transmitter and receiver have antenna arrays. In this case, each measurement can be written as:

$$y_{1 \times 1} = |\mathbf{a}_{1 \times N}^{rx} \mathbf{F}' \mathbf{x}_{N \times 1}^{rx} \mathbf{x}_{1 \times N}^{tx} \mathbf{F}' \mathbf{a}_{N \times 1}^{tx}| \quad (2)$$

where \mathbf{a}^{rx} and \mathbf{a}^{tx} are vectors corresponding to a setting of the phase shifters, \mathbf{F}' is an $N \times N$ inverse Fourier transform

matrix and \mathbf{x}^{rx} and \mathbf{x}^{tx} are sparse vectors representing the angle of arrival at the receiver and the angle of departure at the transmitter respectively. Our goal is to recover \mathbf{x}^{rx} and \mathbf{x}^{tx} from several measurements y .

We can reduce this problem to the problem solved earlier. Specifically, we make $B \times B$ measurements of the form:

$$\mathbf{Y}_{B \times B} = |\mathbf{A}_{B \times N}^{rx} \mathbf{F}' \mathbf{x}_{N \times 1}^{rx} \mathbf{x}_{1 \times N}^{tx} \mathbf{F}' \mathbf{A}_{N \times B}^{tx}| \quad (3)$$

where \mathbf{A}^{rx} is the phase shift matrix as in Theorem 4.1, and \mathbf{A}^{tx} is its transpose. Let \mathbf{A}_i^{rx} be the i -th column of \mathbf{A}^{rx} and \mathbf{A}_j^{tx} be the j -th row of \mathbf{A}^{tx} . Furthermore, let $y_i = \sum_j Y_{i,j}$. We observe that:

$$\begin{aligned} y_i &= \sum_j Y_{i,j} = \sum_j |\mathbf{A}_i^{rx} \mathbf{F}' \mathbf{x}^{rx} \mathbf{x}^{tx} \mathbf{F}' \mathbf{A}_j^{tx}| \\ &= \sum_j |\mathbf{A}_i^{rx} \mathbf{F}' \mathbf{x}^{rx}| |\mathbf{x}^{tx} \mathbf{F}' \mathbf{A}_j^{tx}| \\ &= |\mathbf{A}_i^{rx} \mathbf{F}' \mathbf{x}^{rx}| \left(\sum_j |\mathbf{x}^{tx} \mathbf{F}' \mathbf{A}_j^{tx}| \right) = |\mathbf{A}_i^{rx} \mathbf{F}' \mathbf{x}^{rx}| C \end{aligned}$$

where C is a constant independent of i . Therefore, we can recreate the measurements of the form needed by Theorem 4.2 from the B^2 measurements provided by the matrix \mathbf{Y} . In this way we can test, for each i , whether x_i^{rx} is non-zero, and each test is correct with probability at least $2/3$. By repeating the process $L = O(\log N)$ times, we can detect each non-zero entry with high probability. The total number of measurements is $B^2 L = O(K^2 \log N)$.⁴

Finally, while we described the algorithm for 1D antenna arrays, the algorithm holds for 2D arrays as well. We simply need to apply the hash function along both dimensions of the array. For an $N \times N$ antenna array, the complexity will be $O(K^2 \log N^2)$ and hence will continue to scale logarithmically with the number of antennas in the array.

5 AGILE-LINK IMPLEMENTATION

We have implemented Agile-Link by designing and building a full-fledged mmWave radio capable of fast beam steering, which is shown in Fig.5. The radio operates in the new 24GHz ISM band and serves as a daughterboard for the USRP software radios. Its physical layer supports a full OFDM stack up to 256 QAM. Our implementation addresses critical system and design issues that are described below.

(a) Heterodyne Architecture: Millimeter Wave hardware is significantly more expensive than GHz hardware. Thus, we leverage a heterodyne architecture where the mmWave

⁴In general, the transmitter and receiver can pick the direction with the highest power as the best direction to align their beams. In special cases, where two paths are detected with the similar power, it might be hard to map which path on the transmitter side corresponds to the which path on the receiver side? This can be addressed by either incurring 4 extra measurements to test the path pairs or by performing the soft voting jointly over the path pairs.

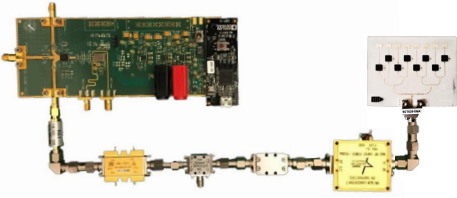


Figure 5: Agile-Link Platform: The figure shows the phased array and mmWave radio we built to operate as a daughterboard for the USRP software radio.

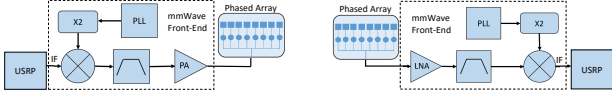


Figure 6: Agile-Link's Architecture. The figure shows block diagrams for both Agile-Link's transmitter and receiver.

signal is first taken into an intermediate frequency of a few GHz, before the I and Q (real & imaginary) components are separated. Such a design reduces the number of components that need to operate at very high frequencies (e.g., mixers, filters, etc) and replaces them with components that operate at a few GHz, which are much cheaper.

The architecture of Agile-Link's receiver is shown in Fig 6. The first block is a mmWave phased array which allow us to steer the beam electronically. The array consists of antenna elements where each element is connected to a phase shifter. The outputs of the phase shifters are combined and fed to a single mmWave front-end. The front-end has a standard design of a low-noise amplifier (LNA), band-pass filter, mixer, and a PLL. The mmWave front-end down-converts the mmWave signal to an intermediate frequency (IF) and feeds it to the daughterboard on the USRP which samples it and passes the digitized samples to the UHD driver. This enables easy manipulation of mmWave signals using standard GNU-radio software and allows us to build an OFDM stack that supports up to 256 QAM.

We have built the design in Fig. 6 using off-the-shelf components. For the mmWave low-noise amplifier (LNA) and power amplifier (PA), we use Hittite HMC-C020 and Quinstar QLW-2440, respectively. For the mmWave mixer, we use Marki M1R-0726MS. To generate local oscillator (LO) signals, we use Analog Devices ADF5355 PLL and Hittite HMC-C035 frequency doubler. The phased array includes 8 antenna elements separated by $\frac{\lambda}{2}$, where each element is connected to a Hittite HMC-933 analog phase shifter. We use Analog Device AD7228 digital-to-analog converters (DAC) and Arduino Due micro-controller board to digitally control the phase shifters.

(b) Radio Performance: To test Agile-Link's ability to deliver high data rates and long range using phased arrays, we measure the SNR of the received signal for different distances

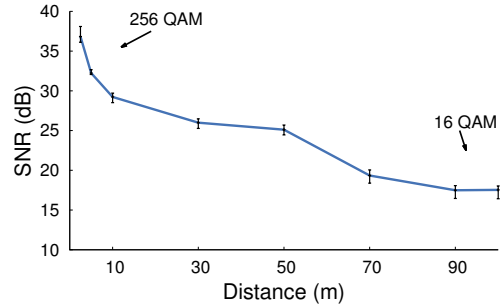


Figure 7: Agile-Link Coverage. SNR at the receiver versus distance between the receiver and the transmitter.

between Agile-Link's receiver and transmitter where the transmit power complies with FCC part15. Fig. 7 shows the SNR at the receiver versus the distance between transmitter and receiver. The figure shows that Agile-Link's implementation provides SNR of more than 30 dB for distances smaller than 10m and 17 dB even at 100m which is sufficient for relatively dense modulations such as 16 QAM [42].

6 EXPERIMENTAL EVALUATION

We evaluate Agile-Link's ability to identify the best beam alignment quickly and accurately. We ran experiments in a lab area with standard furniture (desks, chairs, computers, etc.). We also ran experiments in an anechoic chamber, where we can accurately measure the ground truth. The anechoic chamber walls are covered with RF absorbers to eliminate multipath and isolate the space from exterior interference. This isolation is necessary to measure the ground truth path traveled by the signal without having RF reflections.

6.1 Compared Schemes

We compare the following three schemes:

- **Exhaustive Search:** In this approach, for each setting of the transmitter's beam direction, the receiver scans all different directions. The combination of transmitter and receiver beams that delivered the maximum power is picked as the direction of the signal.
- **802.11ad Standard:** The standard has three phases[22]. The first stage is called *Sector Level Sweep (SLS)*. In this stage, the AP transmits in all possible directions, and the client sets its receiver beam pattern to a quasi-omnidirectional beam. The process is then repeated with the AP setting its receiver antenna to quasi-omnidirectional and the client sweeping through all transmit directions. At the end of this stage, the AP and client each pick the γ directions that deliver the largest power. Note that while some proposals amend the standard to perform SLS with hierarchical beams, hierarchical search is not robust to multipath as described in §3. The second stage is called *Multiple sector ID Detection (MID)*. This

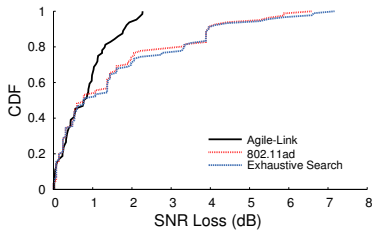


Figure 8: Beam Accuracy with a Single Path: SNR loss due to beam misalignment for Agile-Link, the 802.11ad standard, and exhaustive search.

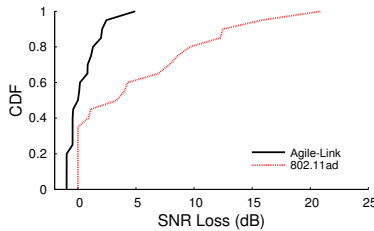


Figure 9: Beam Accuracy with Multi-path: SNR loss of the 802.11ad standard and Agile-Link with respect to the exhaustive search.

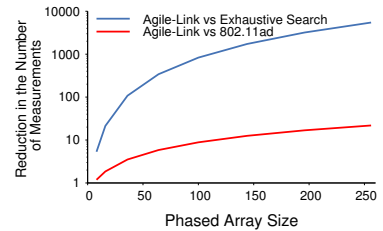


Figure 10: Beam Alignment Latency: Reduction in the number of measurements for Agile-Link versus the 802.11ad standard and exhaustive search.

stage repeats the process above but with the transmit beam set to quasi-omni-directional and the scan being performed with the receive beam. This stage compensates for imperfections in the quasi omnidirectional beams. The third stage is called *Beam Combining (BC)*. In this stage, each of the γ best directions at the AP are tried with each of the γ directions at the client. Hence, γ^2 combinations are tested and the combination of transmit and receive beams that deliver the maximum power is selected and used for beamforming during data transmission. In our experiments, we set $\gamma = 4$.

- **Agile-Link:** We run the algorithm described in §4.2. We set K to 4 since past measurement studies show that the mmWave channel has only 2 to 3 paths [6, 34, 39, 40].

6.2 Beam Alignment Accuracy vs. the Ground Truth

We first evaluate whether Agile-Link can detect the best alignment of the receiver’s and transmitter’s beams. To perform this evaluation, we need to know the ground truth, i.e., the actual direction of the signal. Thus, we run this experiment in an anechoic chamber, where we can accurately measure the ground truth. For each experiment, we place Agile-Link’s transmitter and receiver at two different locations. We then change the orientation of the transmitter’s and receiver’s antenna arrays with respect to each other for all angles between 50° and 130° with increments of 10° . Since there is only a single line-of-sight path in the anechoic chamber, this path will appear at a different direction at the transmitter and at the receiver depending on the orientation of the antenna arrays. Hence, this allows us to test any combination of directions from which the strongest path can leave the transmitter and arrive at the receiver. For each setting, the transmitter transmits measurement frames (as required in 802.11ad) which the receiver uses to compute the directions of the best beam alignment. We then steer the beams based on the output of the alignment and measure the SNR achieved by this alignment.

We will use the SNR loss in comparison to the optimal alignment, i.e., how much SNR could we have gained had we known the ground truth as our accuracy meter. We calculate this metric by measuring the SNR achieved by our beam alignment and subtract it from the SNR achieved by the optimal alignment: $SNR_{loss} = SNR_{optimal} - SNR_{Agile-Link}$. The lower the SNR loss, the higher is our accuracy in detecting the direction of the signal.

Fig. 8 plots a CDF of the SNR loss for Agile-Link’s beam searching scheme, the exhaustive search and the 802.11ad standard, in comparison to the optimal alignment. The figure reveals two interesting points. First, it shows that Agile-Link performs better than the two baselines in that it has minimal SNR loss. While all schemes have a median SNR loss below 1dB, the 90^{th} percentile SNR loss for both exhaustive search and the standard is 3.95dB which is higher than the 1.89dB SNR loss of Agile-Link. This is due to the fact that the standard and exhaustive search choose to steer using the best beam from a discrete set of N beams which they tested. However, the space of beam directions is continuous and the best beam may not exactly align with the discretization chosen by the algorithms. In this case, they will end up picking the closest beam in the discrete set, which may not be the exact optimal one. SNR loss is further exacerbated by the fact that this can happen on both sides i.e., the transmitter and the receiver. In contrast, Agile-Link uses the beams as a continuous weight over the possible choice of directions (Equation 1) and picks the direction that maximizes the overall weight, as described in §4.2. Thus, Agile-Link can discover the direction of the path beyond the N directions used by exhaustive search and the standard.

Furthermore, the figure shows that the standard and exhaustive search have similar performance. This might seem surprising since one may expect exhaustive search to find a better beam alignment since it spends more time searching the space. However, it is important to recall that the standard differs from the exhaustive search only in the first stage where it uses a quasi-omnidirectional beams to limit the search space to a few top candidates. In the final the

stage, however, the standard tries all possible combinations of these candidate beams. Since there is only one path in this experiment, as long as the best beam is picked as one of the candidate beams in the first stage, the standard will converge to the same beam alignment as the exhaustive search. However, we will show next that this does not continue to hold in multipath settings.

6.3 Alignment Accuracy in Multipath Environments

We repeat the above experiments in an office environment, where due to multipath, the signal can arrive from different directions. However, in this case, we do not know the ground truth for the direction of strongest path and hence we measure the SNR loss relative to the exhaustive search baseline: $SNR_{loss} = SNR_{Exhaustive} - SNR_{Agile-Link}$. Note that since exhaustive search tries all possible combinations of directions, it maintains its performance with multipath.

Fig. 9 plots a CDF of SNR loss for Agile-Link and the 802.11ad standard with respect to the exhaustive search. The figure shows that the standard performs much worse in multipath scenarios. Specifically, instead of having a similar SNR to the exhaustive search as before, the median and 90th percentile SNR loss (with respect to exhaustive search) are 4dB and 12.5dB, respectively. This is because the standard is using its phased array as a quasi-omnidirectional antenna and hence the multiple paths can combine destructively, in which case the information is lost. Further, due to imperfections in the quasi-omnidirectional patterns, some paths can get attenuated and hence the standard can easily choose the wrong direction to align its beam. In contrast, Agile-Link performs well even in the presence of multipath. Specifically, the median and 90th percentile SNR loss with respect to exhaustive search are 0.1dB and 2.4dB, respectively. Finally, the figure also shows that sometimes Agile-Link’s SNR loss with respect to exhaustive search is negative. This is because in some cases, Agile-Link performs better than exhaustive search for the same reasons described above.

6.4 Beam Alignment Latency

Next we would like to evaluate the gain in reducing latency that Agile-Link delivers over the two baselines. However, since our radio has a fixed array size we cannot empirically measure how this gain scales for larger arrays. Hence, we perform extensive simulations to compute this gain for larger arrays and we use the empirical results from our 8-antenna array to find the delay for this array size.

(a) *Reduction in the Number of Measurements:* Since each measurement in 802.11ad requires sending a special frame, one way to measure delay is in terms of the number of measurements frames. Fig. 10 plots the reduction in the number

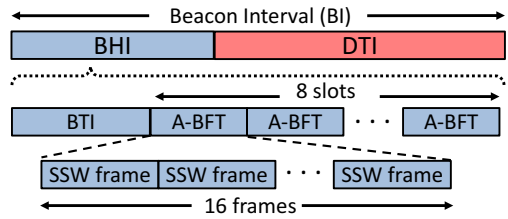


Figure 11: 802.11ad Beacon Interval Structure

of measurements that Agile-Link achieves over exhaustive search and the standard. For an 8-antenna phased array, Agile-Link reduces the number of measurements by 7× and 1.5× compared to exhaustive search and standard, respectively. Further, the gain increases quickly as the number of antennas increase. This is due to the scaling property of each algorithm and whether it is quadratic, linear, or logarithmic. For arrays of size 256, Agile-Link is 16.4× better than the standard and three orders of magnitude better than exhaustive search.

(b) *Reduction in Search Time:* Next, we look at the amount of time it takes to find the best alignment in each scheme under the 802.11ad MAC protocol. The standard is still evolving; our description is based on [22]. Since the delays in exhaustive search are unacceptable in practice, we consider only Agile-Link and the beam alignment scheme in the standard.

The 802.11ad has a protocol for when the AP and clients search the space to align their beams [22, 28]. The delay produced through this process differs from simply multiplying the number of measurement frames by the duration of each measurement. This is due to three main reasons: 1) The protocol allows for beam scan (called beam training) only during certain intervals. If the client cannot collect all necessary measurements, it needs to wait until the next opportunity to perform more measurements. 2) Different clients contend for the beam alignment slots; hence, the delay will increase depending on the number of clients. 3) When the AP sweeps its beam, all clients can collect measurements; hence this part can be amortized.

To better understand the above constraints, let us describe at a high-level how 802.11ad performs beam-forming training. The AP periodically transmit beacon intervals (BI), which typically last for 100 ms [28]. Each BI has a beacon header intervals (BHI), followed by a data transmission interval (DTI), as shown in Fig. 11. The search for the best alignment is done during the BHI. Each BHI consists of one beacon transmission interval (BTI) which is used by the AP to train its antenna beam, and eight association beam-forming training (A-BFT) slots, which are randomly selected by clients to train their beams. Finally, each A-BFT slot consists of up to 16 SSW frames, where each frame is used to perform one measurement and has a duration of 15.8μs [3, 22]. Each BI

Size	One Client		Four Clients	
	802.11ad	Agile-Link	802.11ad	Agile-Link
8	0.51ms	0.44ms	1.27ms	1.20ms
16	1.01ms	0.51ms	2.53ms	1.26ms
64	4.04ms	0.89ms	304.04ms	2.40ms
128	106.07ms	0.95ms	706.07ms	2.46ms
256	310.11ms	1.01ms	1510.11ms	2.53ms

Table 1: Beam alignment latency for different array size

has a maximum of 8 beam training slots. All clients contend for training in those slots. If the client cannot finish its training during one A-BFT, it can contend for further slots during the same or following BI. Yet, waiting for the next BI increases the delay by 100ms.

As explained in §6.1, 802.11 performs beam refinement where each of γ best directions at the AP and client are compared again. To simplify the computation, we conservatively ignore the 802.11ad beam refinement since it only increases the delay of 802.11ad, and improves the relative gains of Agile-Link. Also when simulating 4 client, we assume that the contention succeeded without collision. This is a conservative assumption since Agile-Link requires significantly fewer measurement slots and hence, given the same number of slots, the collision probability between clients is smaller in Agile-Link. Finally, the AP trains its beam during the BTI, and uses frames similar to those used for the client beam training. The AP doesn't need to repeat this training per client.

Table 1 shows the beam alignment delay for different antenna-array sizes, for the case of one client and 4 clients. As the number of antennas in the array increases, the delay in 802.11ad increases quickly. In contrast, Agile-Link can operate within the same standard, but it extracts more information from each measurement, hence keeping the delay low even for large antenna arrays. In particular, for antenna arrays of 256 elements, the proposed 802.11 beam alignment algorithm takes hundreds of milliseconds for one user and over 1.5 seconds for 4 users. In contrast, Agile-Link keeps the delay below 1.01ms and 2.53ms, respectively.

6.5 Comparison with Beam Alignment using Compressive Sensing

Concurrent to our work, the authors of [35] proposed a scheme that leverages compressive sensing to speed up the search for the best beam alignment. Here, we compare Agile-Link with that scheme. On the theoretical front, both algorithms exploit sparsity to speed up the search. However, Agile-Link provably finds the best alignment in a logarithmic number of measurements, while [35] does not provide any theoretical guarantees (note that the standard compressive sensing analysis does not apply because the algorithm of [35] only uses the magnitudes of the measurements). In

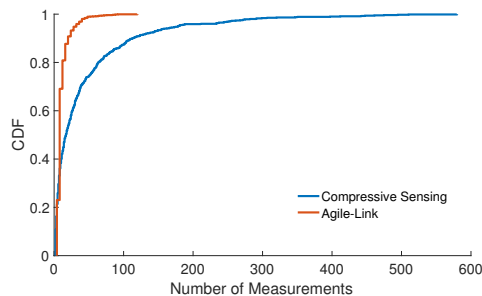


Figure 12: Agile-Link versus Compressive Sensing: Required number of measurements to find the correct direction for Agile-Link versus Compressive Sensing [35].

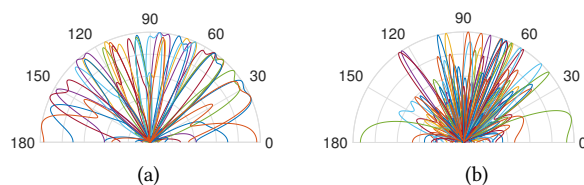


Figure 13: Hashing Beam Patterns Beam patterns of the first 16 measurements for (a) Agile-Link and (b) Compressive Sensing Beam Alignment.

the following, we compare the two schemes empirically to better understand the difference in their performance.

To be fair to both schemes, we use the same channel information from our measurements and run trace driven simulations for antenna arrays of 16 elements. This guarantees that we apply both schemes on the same set of channels. We fix the signal direction of the transmitter. The receiver tries both schemes (one after the other) until it finds the optimal beam alignment. We repeat the experiment 900 times for different channel values, where the channels are taken from empirical measurements in our testbed. Fig. 12 plots the required number of measurements to find the correct direction for both Agile-Link and the compressive sensing scheme. As before, the required number of measurements is defined as the number of measurements until the resulting beam power is within 3dB of the correct optimal beam power. The figure shows that Agile-Link requires significantly fewer measurements. Specifically, Agile-Link needs a median of 8 measurements and a 90th percentile of 20 measurements. In contrast, the compressive sensing scheme needs a median of 18 measurements and a 90th percentile of 115 measurements. Interestingly, the figure shows that the tail of the compressive sensing scheme is pretty high.

To understand the root cause of this behavior, let us look at the beam patterns generated by both schemes. Fig. 13 plots the beam patterns of the first 16 measurements for both schemes. Each color refers to a different multi-armed

beam.⁵ The figure shows that beam shape of the compressive sensing scheme is quite random and hence fails to sample the space uniformly. Further, the figure shows the combination of the first 16 measurements. For Agile-Link it is clear that the first 16 measurements span the space well and hence minimize the probability of missing the right signal direction. In contrast, the first 16 measurements from the compressive sensing scheme do not manage to span the space and leave many signal directions uncovered. Therefore, the compressive sensing scheme has a fairly long tail.

7 CONCLUSION

This paper presents Agile-Link, a phased array mmWave system that can find the best beam alignment without scanning the entire space. Agile-Link delivers a mmWave beam alignment algorithm with provably logarithmic measurements for the phased-array architecture commonly used in mmWave access points and clients. It finds the correct alignment of the beams between a transmitter and a receiver orders of magnitude faster than existing radios. We believe Agile-Link brings us closer towards practical mmWave networks.

ACKNOWLEDGEMENT

This work is supported by NSF. We thank the reviewers and our shepherd Marco Gruteser for their feedback and insightful comments.

APPENDIX

A FORMAL PROOF OF THE ALGORITHM

A.1 Notation and Preliminaries

(a) Basic notation

- We use $[N]$ to denote $\{0 \dots N - 1\}$.
- We use S to denote the support of \mathbf{x} .
- We use \mathbf{F} to denote the Fourier transform matrix, and \mathbf{F}' to denote the inverse Fourier transform matrix. Also, we use \mathbf{F}_i to denote the i -th row of \mathbf{F} ; same for \mathbf{F}' . Finally, we use $\hat{\mathbf{x}}$ to denote $\mathbf{F}\mathbf{x}$.
- For two vectors \mathbf{x} and \mathbf{y} , we define the *Hadamard product* \circ of \mathbf{x} and \mathbf{y} as $(\mathbf{x} \circ \mathbf{y})_i = x_i y_i$. We will use this notion to mask out the coefficients outside of a given segment.
- We use $\mathbf{x} * \mathbf{y}$ to denote the convolution of \mathbf{x} and \mathbf{y} .
- The vectors $\mathbf{e}_0 \dots \mathbf{e}_{N-1}$ denote the standard basis. I.e., $(\mathbf{e}_p)_i = 1$ for $p = i$ and $(\mathbf{e}_p)_i = 0$ otherwise.

(b) Measurements and box car filter

Our measurements can be described using the notion of the

⁵Note these are the actual beams from substituting in the equations for both schemes, while the beams in Fig. 4 are ideal beams. Agile-Link's beams are pretty close to the ideal beams whereas compressive sensing beams deviate strongly from that ideal pattern.

boxcar filter, defined as follows. For parameter P , let \mathbf{H} be such that $H_i = \frac{\sqrt{N}}{P-1}$ if $|i| < P/2$ and $H_i = 0$ otherwise. It is known that $\hat{H}_j = \frac{\sin(\pi(P-1)j/N)}{(P-1)\sin(\pi j/N)}$.

PROPOSITION A.1. $\hat{\mathbf{H}}$ satisfies the following properties: (i) $\hat{H}_0 = 1$; (ii) $\hat{H}_j \in [\frac{1}{2\pi}, 1]$ for $|j| \leq \frac{N}{2P}$; (iii) $|\hat{H}_j| \leq \frac{2}{1+|j|P/N}$ if $P \geq 3$.

CLAIM A.2.

$$\|\hat{\mathbf{H}}\|_2^2 = \sum_j |\hat{H}_j|^2 \leq 1 + 2N/P \sum_j 1/|j|^2 \leq C \frac{N}{P}$$

for some constant C .

We also define a *shifted* version of \mathbf{H} defined as $(H^t)_i = H_{i-t}$. By the time-shift theorem it follows that $|\hat{H}_i^t| = \hat{H}_i$.

Using this notation, we can write each measurement \mathbf{a}^b as

$$\mathbf{a}^b = \sum_{r=0}^{R-1} (\mathbf{F}_{s_b^r} \circ \mathbf{H}^{rN/R}) \omega^{t_r^b}$$

Each segment of \mathbf{a}^b , when multiplied by a row of the matrix \mathbf{F}' , can be interpreted as follows.

$$\text{CLAIM A.3. } (\mathbf{F}_i \circ \mathbf{H}) \cdot \mathbf{F}'_p = \hat{H}_{i-p}$$

PROOF.

$$(\mathbf{F}_i \circ \mathbf{H}) \cdot \mathbf{F}'_p = \mathbf{F}_i \cdot (\mathbf{F}'_p \circ \mathbf{H}) = (\widehat{\mathbf{F}'_p \circ \mathbf{H}})_i = (\mathbf{F}'_p * \hat{\mathbf{H}})_i = (\mathbf{e}_p * \hat{\mathbf{H}})_i = \hat{H}_{i-p}$$

□

(c) Pseudo-random permutations

We will use matrices $\mathbf{P}_{\rho,b}$ parameterized by mappings ρ of the form $\rho(i) = \sigma^{-1}i + a \pmod{N}$ for $\sigma, a, b \in [N]$ such that

- $(\mathbf{P}_{\rho,b}\mathbf{x})_{\rho(i)} = x_i \omega^{\tau(j)}$ for $\tau(j) = b(j + \sigma a)$
- $\mathbf{F}'\mathbf{P}_{\rho,b} = \mathbf{P}'_{\rho,b}\mathbf{F}$ for $\mathbf{P}'_{\rho,b}$ as defined in the paper.

Note that τ is a permutation assuming b is invertible mod N . We use \mathcal{R} to denote the set of all mappings ρ as defined above. For the analysis, we will assume that N is prime. This will ensure that the elements $\rho \in \mathcal{R}$ are permutations. Furthermore, in this case \mathcal{R} is *pairwise independent*, i.e., for any $i \neq j, i' \neq j'$, we have

$$Pr_{\rho \in \mathcal{R}}[\rho(i) = i', \rho(j) = j'] = 1/N^2$$

It will be convenient to assume $\|\mathbf{x}\|_2^2 = 1$. Then we can define the threshold T to be $(\frac{1}{2(2\pi)} - \frac{1}{8\pi})^2 (\frac{1}{2(2\pi)})^2 / K$.

A.2 Proofs

LEMMA A.4. Fix b and select $\rho \in \mathcal{R}$ uniformly at random. Then, for any s :

$$E[I(b, \rho(s))] = E[|\mathbf{a}^b \mathbf{F}'_{\rho(s)}|^2] \leq CR/P$$

where C is the constant from Claim A.2.

PROOF.

$$E \left[|\mathbf{a}^b \mathbf{F}'_{\rho(s)}|^2 \right] = E \left[\left| \sum_{r=0}^{R-1} ((\mathbf{F}_{s_b^r} \circ \mathbf{H}^{rN/R}) \omega^{t_r^b}) \mathbf{F}'_{\rho(s)} \right|^2 \right] \quad (4)$$

$$= E \left[\left| \sum_{r=0}^{R-1} \hat{H}_{s_b^r - \rho(s)}^{rN/R} \omega^{t_r^b} \right|^2 \right] \quad (5)$$

$$= \sum_{r=0}^{R-1} E \left[|\hat{H}_{s_b^r - \rho(s)}|^2 \right] \quad (6)$$

where in step 5 we used Claim A.3 and the independence of the variables t_r^b , $r = 0 \dots R-1$. Since $i = s_b^r - \rho(s)$ is distributed uniformly at random in $[N]$, by Claim A.2:

$$\sum_{r=0}^{R-1} E \left[|\hat{H}_{s_b^r - \rho(s)}|^2 \right] \leq \sum_{r=0}^{R-1} 1/N \sum_i |\hat{H}_i|^2 \leq R/N \cdot N/P \cdot C = CR/P$$

□

LEMMA A.5. *Suppose that $|s_b^r - i| \leq \frac{N}{2P}$. Then*

$$\Pr[|\mathbf{a}^b \mathbf{F}'_i|^2 \geq \frac{1}{4(2\pi)^2}] \geq 5/6$$

PROOF.

$$|\mathbf{a}^b \mathbf{F}'_i|^2 = \left| \sum_{r'=0}^{R-1} \hat{H}_{s_b^{r'} - i}^{r'N/R} \omega^{t_{r'}^b} \right|^2 \quad (7)$$

$$= |\hat{H}_{s_b^r - i} \omega^{t_r^b} + \sum_{r' \neq r} \hat{H}_{s_b^{r'} - i} \omega^{t_{r'}^b}|^2 \quad (8)$$

$$= |\hat{H}_{s_b^r - i} \omega^{t_r^b} - X|^2 \quad (9)$$

We know from Proposition A.1 that $|\hat{H}_{s_b^r - i} \omega^{t_r^b}| \geq \frac{1}{2\pi}$. We will show that the probability of $X \geq \frac{1}{2(2\pi)}$ is at most 1/6. It will follow that $|\mathbf{a}^b \mathbf{F}'_i|^2 \geq \frac{1}{4(2\pi)^2}$ with the probability of at least 5/6.

Recall that $s_b^0, s_b^1 \dots$ are separated by P . Therefore, for $r' \neq r$, we have $|s_b^{r'} - i| \geq P - |s_b^r - i| \geq P - \frac{N}{2P}$, which is at least $P/2$ for B large enough. By the independence of the variables $t_{r'}^b$ and by Proposition A.1 we have:

$$E[X^2] = E \left[\left| \sum_{r' \neq r} \hat{H}_{s_b^{r'} - i} \omega^{t_{r'}^b} \right|^2 \right]$$

$$= \sum_{r' \neq r} E \left[|\hat{H}_{s_b^{r'} - i} \omega^{t_{r'}^b}|^2 \right]$$

$$\leq 2 \sum_{d=1}^R \left(\frac{2}{1 + P/N \cdot d(P/2)} \right)^2$$

$$\leq 8/(P^2/N) \sum_{d=1}^R (1/d)^2 \leq 8CN/P^2$$

Since $P = N/R$ and $R^2 = N/B$, the latter expression is bounded by $8C/B$, which is less than $\frac{1}{6(4\pi)^2}$ for B large enough. It follows that the probability of $X^2 \geq \frac{1}{(4\pi)^2} = 6E[X^2]$ is at most 1/6. □

PROOF OF THEOREM 4.1, PART (1). Suppose that $s \in S$. Select s_b^r that is closest to $\rho(s)$. Note that $|s_b^r - \rho(s)| \leq \frac{N}{2P}$, which by Lemma A.5 implies $|\mathbf{a}^b \mathbf{F}'_{\rho(s)}|^2 \geq \frac{1}{4(2\pi)^2}$ with the probability of at least 5/6.

We now lower bound $T(s)$ as follows

$$T(s) \geq |\mathbf{a}^b \mathbf{F}'_{\rho, b\mathbf{x}}|^2 |\mathbf{a}^b \mathbf{F}'_{\rho, b\mathbf{e}_s}|^2 \quad (10)$$

$$= |Y\mathbf{x}_s - X|^2 |Y|^2 \quad (11)$$

where $Y = \omega^{\tau(s)} \mathbf{a}^b \mathbf{F}'_{\rho(s)}$ and $X = \sum_{s' \in S - \{s\}} \omega^{\tau(s')} \mathbf{a}^b \mathbf{F}'_{\rho(s')} \mathbf{x}_{s'}$.

We can bound $E[|X|^2]$ as follows.

$$E \left[\left| \sum_{s' \in S - \{s\}} \omega^{\tau(s')} \mathbf{x}_{s'} \mathbf{a}^b \mathbf{F}'_{\rho(s')} \right|^2 \right] \quad (12)$$

$$= \sum_{s' \in S - \{s\}} \mathbf{x}_{s'}^2 E[|\mathbf{a}^b \mathbf{F}'_{\rho(s')}|^2] \quad (13)$$

$$\leq \sum_{s' \in S - \{s\}} \mathbf{x}_{s'}^2 CR/P \quad (14)$$

$$= \|\mathbf{x}\|_2^2 CR/P \leq C/B \leq 1/K \frac{1}{6 \cdot (8\pi^2)^2} \quad (15)$$

where we used Parseval's identity, Lemma A.4 and that B is large enough. Therefore, we have that $\Pr[X^2 \geq \frac{1}{(8\pi)^2}] \leq 1/6$. By Lemma A.5 we have that, with probability at least $1 - 1/6 - 1/6$, $T(s) \geq (\frac{1}{2(2\pi)} - \frac{1}{8\pi})^2 (\frac{1}{2(2\pi)})^2 / K$. □

PROOF OF THEOREM 4.1, PART (2). Suppose that $s \notin S$. We have

$$E[T(s)]$$

$$\leq \sum_{b=0}^{B-1} E_{\rho(s), \rho(s'), \tau} \left[\left| \sum_{s' \in S} \omega^{\tau(s')} \mathbf{x}_{s'}^2 \mathbf{a}^b \mathbf{F}'_{\rho(s')} \right|^2 |\mathbf{a}^b \mathbf{F}'_{\rho(s)}|^2 \right]$$

$$= \sum_{b=0}^{B-1} E_{\rho(s'), \tau} \left[\left| \sum_{s' \in S} \omega^{\tau(s')} \mathbf{x}_{s'}^2 \mathbf{a}^b \mathbf{F}'_{\rho(s')} \right|^2 \right] E_{\rho(s)} \left[|\mathbf{a}^b \mathbf{F}'_{\rho(s)}|^2 \right]$$

$$\leq CR/P \sum_{b=0}^{B-1} \sum_{s' \in S} \mathbf{x}_{s'}^2 E[|\mathbf{a}^b \mathbf{F}'_{\rho(s')}|^2]$$

$$\leq (CR/P)^2 B \|\mathbf{x}\|_2^2 \leq C^2/B \leq T/3$$

where we assume that B is large enough. By Markov inequality it follows that $\Pr[T(s) \geq T] \leq 1/3$. □

REFERENCES

- [1] [n. d.]. *FCC to explore 5G services above 24 GHz*. <http://www.fiercewireless.com>.
- [2] 2015. 5G Vision White Paper - Samsung. (2015).

- [3] 2016. Short SSW Format for 11ay. <https://mentor.ieee.org/802.11/dcn/16/11-16-0416-01-00ay-short-ssw-format-for-11ay.pptx>. (2016).
- [4] 60 GHz: Taking the VR Experience to the Next Level. [n. d.]. <http://www.sibeam.com>. ([n. d.]).
- [5] Ahmed Alkhateeb, Omar El Ayach, Geert Leus, and Robert W Heath. 2014. Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE Selected Topics in Signal Processing* (2014).
- [6] Christopher R. Anderson and Theodore S. Rappaport. 2004. In-Building Wideband Partition Loss Measurements at 2.5 and 60 GHz. *IEEE TWC* (2004).
- [7] Khanh Do Ba, Piotr Indyk, Eric Price, and David P Woodruff. 2010. Lower bounds for sparse recovery. In *SODA*.
- [8] Matt Branda. 2015. Qualcomm Research demonstrates robust mmWave design for 5G. Qualcomm Technologies Inc.. (November 2015).
- [9] E. Candes, J. Romberg, and T. Tao. 2006. Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information. *IEEE Transactions on Information Theory* (2006).
- [10] Cisco. 2013. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update. (2013).
- [11] D. Donoho. 2006. Compressed Sensing. *IEEE Trans on Info. Theory* (2006).
- [12] Mohammed E. Eltayeb, Ahmed Alkhateeb, Robert W. Heath, and Tareq Y. Al-Naffouri. 2015. Opportunistic Beam Training with Hybrid Analog/Digital Codebooks for mmWave Systems. In *GLOBESIP*.
- [13] Bo Gao, Zhenyu Xiao, Changming Zhang, Depeng Jin, and Lieguang Zeng. 2014. Sparse/dense channel estimation with non-zero tap detection for 60-GHz beam training. *IET Communications* 8, 11 (2014).
- [14] A. Gilbert, M. Muthukrishnan, and M. Strauss. 2005. Improved time bounds for near-optimal space Fourier representations. In *SPIE*.
- [15] Anna C Gilbert, Sudipto Guha, Piotr Indyk, S Muthukrishnan, and Martin Strauss. [n. d.]. Near-optimal sparse Fourier representations via sampling. In *STOC'02*.
- [16] Muhammad K. Haider and Edward W. Knightly. 2016. Mobility resilience and overhead constrained adaptation in directional 60 GHz WLANs: protocol design and system implementation. In *MobiHoc*.
- [17] Daniel Halperin, Srikanth Kandula, Jitendra Padhye, Paramvir Bahl, and David Wetherall. 2011. Augmenting Data Center Networks with Multi-Gigabit Wireless Links. In *ACM SIGCOMM*.
- [18] Haitham Hassanieh, Piotr Indyk, Dina Katabi, and Eric Price. 2012. Nearly optimal sparse fourier transform. In *STOC*.
- [19] Haitham Hassanieh, Piotr Indyk, Dina Katabi, and Eric Price. 2012. Simple and practical algorithm for sparse FFT. In *SODA*.
- [20] Ken'ichi Hosoya, Narayan Prasad, Kishore Ramachandran, Naoyuki Orihashi, Shuya Kishimoto, Sampath Rangarajan, and Kenichi Maruhashi. 2015. Multiple sector ID capture (MIDC): A novel beamforming technique for 60-GHz band multi-Gbps WLAN/PAN systems. *IEEE Transactions on Antennas and Propagation* (2015).
- [21] IBM Breakthrough Could Alleviate Mobile Data Bottleneck. [n. d.]. <http://cacm.acm.org/>. ([n. d.]).
- [22] IEEE Standards Association. 2012. IEEE Standards 802.11ad-2012: Enhancements for Very High Throughput in the 60 GHz Band. (2012).
- [23] Mark Iwen, Aditya Viswanathan, and Yang Wang. 2017. Robust sparse phase retrieval made easy. *Applied and Computational Harmonic Analysis* (2017).
- [24] Kishore Jaganathan, Yonina C Eldar, and Babak Hassibi. 2015. Phase retrieval: An overview of recent developments. *arXiv* (2015).
- [25] John Kilpatrick, Robbie Shergill, and Manish Sinha. [n. d.]. 60 GHz Line of Sight Backhaul Links Ready to Boost Cellular Capacity. Analog Devices. ([n. d.]).
- [26] Bin Li, Zheng Zhou, Weixia Zou, Xuebin Sun, and Guanglong Du. 2013. On the Efficient Beam-Forming Training for 60GHz Wireless Personal Area Networks. *IEEE Transactions on Wireless Communications* 12, 2 (February 2013).
- [27] Thomas Nitsche, Guillermo Bielsa, Irene Tejado, Adrian Loch, and Joerg Widmer. 2015. Boon and bane of 60 GHz networks: Practical insights into beamforming, interference, and frame level operation. In *CoNEXT*.
- [28] Thomas Nitsche, Carlos Cordeiro, Adriana B Flores, Edward W Knightly, Eldad Perahia, and Joerg C Widmer. 2014. IEEE 802.11 ad: directional 60 GHz communication for multi-Gigabit-per-second Wi-Fi. *IEEE Comm. Magazine* (2014).
- [29] Thomas Nitsche, Adriana B Flores, Edward W Knightly, and Joerg Widmer. 2015. Steering with eyes closed: mm-wave beam steering without in-band measurement. In *INFOCOM*.
- [30] Zhouyue Pi and Farooq Khan. 2011. An introduction to millimeter-wave mobile broadband systems. *Communications Magazine, IEEE* (2011).
- [31] Eric Price and David P Woodruff. 2011. (1+ eps)-approximate sparse recovery. In *FOCS*.
- [32] Hariharan Rahul, Swarun Suresh Kumar, and Dina Katabi. 2012. MegaMIMO: Scaling Wireless Capacity with User Demand. In *SIGCOMM*.
- [33] Dinesh Ramasamy, Subramanian Venkateswaran, and Upamanyu Madhow. 2012. Compressive tracking with 1000-element arrays: A framework for multi-Gbps mm wave cellular downlinks. In *Allerton*.
- [34] Sundeep Rangan, Theodore S Rappaport, and Elza Erkip. 2014. Millimeter-wave cellular wireless networks: Potentials and challenges. *IEEE* (2014).
- [35] Maryam Eslami Rasekh, Zhinus Marzi, Yanzi Zhu, Upamanyu Madhow, and Haitao Zheng. 2017. Noncoherent mmWave path tracking. In *ACM HotMobile*.
- [36] Wonil Roh, Ji-Yun Seol, JeongHo Park, Byunghwan Lee, Jaekon Lee, Yungsoo Kim, Jaeweon Cho, Kyungwhoon Cheun, and Farshid Aryanfar. 2014. Millimeter-Wave Beamforming as an Enabling Technology for 5G Cellular Communications: Theoretical Feasibility and Prototype Results. *IEEE Communications Magazine* (2014).
- [37] Clayton Shepard, Hang Yu, Narendra Anand, Li Erran, Thomas Marzetta, Richard Yang, and Lin Zhong. [n. d.]. Argos Practical Many-Antenna BaseStations. In *MobiCom'12*.
- [38] SiBeam, Lattice Semiconductor. [n. d.]. www.sibeam.com. ([n. d.]).
- [39] Sanjib Sur, Vignesh Venkateswaran, Xinyu Zhang, and Parameswaran Ramanathan. 2015. 60 GHz Indoor Networking through Flexible Beams: A Link-Level Profiling. In *SIGMETRICS*.
- [40] Sanjib Sur, Xinyu Zhang, Parameswaran Ramanathan, and Ranveer Chandra. 2016. BeamSpy: Enabling Robust 60 GHz Links Under Blockage. In *NSDI*.
- [41] Y. Ming Tsang, Ada S. Y. Poon, and Sateesh Addepalli. 2011. Coding the Beams: Improving Beamforming Training in mmWave Communication System. In *IEEE GLOBECOM*.
- [42] D. Tse and P. Vishwanath. 2005. *Fundamentals of Wireless Communications*. Cambridge University Press.
- [43] UMTS Forum. 2011. Mobile Traffic Forecasts: 2010-2020 Report. (2011).
- [44] Jie Xiong and Kyle Jamieson. 2013. ArrayTrack: A Fine-Grained Indoor Location System. In *NSDI*.
- [45] Wenfang Yuan, Simon Armour, and Angela Doufexi. [n. d.]. An Efficient and Low-complexity Beam Training Technique for mmWave Communication. In *PIMRC'15*.
- [46] Liang Zhou and Yoji Ohashi. 2012. Efficient Codebook-Based MIMO Beamforming for Millimeter-Wave WLANs. In *PIMRC*.
- [47] Yibo Zhu, Zengbin Zhang, Zhinus Marzi, Chris Nelson, Upamanyu Madhow, Ben Y. Zhao, and Haitao Zheng. [n. d.]. Demystifying 60GHz Outdoor Picocells. In *MOBICOM'14*.